

昭62-255999

⑤Int.Cl.

G 10 L 3/00

識別記号

301

庁内整理番号

D-8221-5D  
E-8221-5D  
B-8221-5D

④公開 昭和62年(1987)11月7日

審査請求 有 発明の数 1 (全 8 頁)

⑤発明の名称 単語音声認識装置

⑥特 願 昭61-98118

⑦出 願 昭61(1986)4月30日

⑧発明者 藤本 敦幸 川崎市中原区上小田中1015番地 富士通株式会社内

⑨出願人 富士通株式会社 川崎市中原区上小田中1015番地

⑩代理人 弁理士 滝野 秀雄 外2名

## 明細書

## 1. 発明の名称

単語音声認識装置

## 2. 特許請求の範囲

入力単語音声パターンを単音節標準パターンから作成された擬似単語標準パターンと照合して入力単語音声を認識する単語音声認識装置において、

(a) 入力単語音声より作成された入力単語音声パターンより無音区間パターンを取り除き、各有音区間パターンを結めて圧縮単語音声パターンを作成する圧縮単語音声パターン作成手段 (110) と、

(b) 各単音節標準パターンより認識対象となるカテゴリの単語群に属する各単語の擬似単語標準パターンを作成する擬似単語標準パターン作成手段 (120) と、

(c) 圧縮単語音声パターン作成手段 110 より入力された圧縮単語音声パターンを擬似単語標準パターン作成手段 (120) 中の各擬似単語標

準パターンと照合して入力単語音声を認識する

単語認識手段 (130) 、

を備えたことを特徴とする単語音声認識装置。

## 3. 発明の詳細な説明

## (概要)

入力単語音声パターンを単音節標準パターンから作成された擬似単語標準パターンと照合する単語音声認識装置において、入力単語音声パターンより無音区間パターンを取り除き、各有音区間パターンを結めて作成された圧縮単語音声パターンを用いて擬似単語標準パターンと照合する。これにより認識率を向上させると共に処理量を低減させることが出来る。

## (産業上の利用分野)

本発明は、単語音声を認識する単語音声認識装置、特に、登録された単音節標準パターンと未知入力単語音声パターンとを照合して入力単語音声を認識する単語音声認識装置において、入力単語

音声パターン中に存在する無音区間パターンによる悪影響を除去して認識率を向上させると共に処理量を低減させる様に改良した単語音声認識装置に関する。

未知入力単語音声を認識する場合、入力単語音声から作成された入力単語音声パターンを予め登録されている単語標準パターンと照合する認識方式が多く用いられている。

この単語音声認識方式において単語標準パターンを登録する場合、実際に発声された単語音声より作成された単語標準パターンを用いる方式と、予め登録されている単音節標準パターンを連結して作成された擬似単語標準パターンを用いる方式がある。

前者の単語標準パターンを用いる方式は、認識率は良好であるが、認識対象となる単語の数だけ単語標準パターンを登録する必要がある為、認識単語数が増加すると、登録作業に多くの手間と時間が掛かり、且つ、認識対象となる単語群のカテゴリが変更されると、再び登録をやり直さねばなら

ないという不都合がある。

これに対し、後者の擬似単語標準パターンを用いる方式は、認識率の点では前者の方式より一般的に劣るが、約100種類程の単音節標準パターンを登録するだけで、任意の擬似単語標準パターンを作成することが可能であり、認識対象となる単語群のカテゴリが変更になっても再登録する必要がないので、登録作業が簡単で済む利点がある。

なお、各単語は音節(シラブル)から成り立ち、音節は音素から成り立っている。音素は音声の最小基本単位で、母音と子音がある。各音節は、通常1個の母音と1ないし2個の子音が結合して形成され、日本語の場合、約100種の音節がある。

本発明は、後者の擬似単語標準パターンを用いる単語音声認識方式に関し、その認識率を向上させる様にしたものである。

#### 〔従来の技術〕

第5図は、従来の単音節標準パターンから作成された擬似単語標準パターンによる単語音声認識

方式の基本構成をブロック図で示したものである。

第5図において、未知の入力単語音声が図示しないマイクロホンから入力されると、音声分折部210は、入力単語音声の特徴を表すパラメタや各音節の区間検出等を行って音節対応の入力単語音声パターンを作成し、単語認識部220に入力する。

一方、単音節標準パターン辞書230には、各単音節標準パターンが予め登録されており、認識対象となる単語群のカテゴリが決ると、単音節標準パターン辞書230から単音節標準パターンを取り出して連結することにより、認識対象カテゴリに属する各単語に対応する擬似単語標準パターンが作成され、擬似単語標準パターン辞書240に格納される。

単語認識部220は、音声分折部210より入力された入力単語音声パターンを擬似単語標準パターン辞書240中の各擬似単語標準パターンと照合し、距離の最も小さい擬似単語標準パターンの単語を認識単語とする。

単語認識部220における、前述の単語認識処理は、2段DP法(Two level dynamic programming matching)によって行われるが、第6図は、そのDPマッチング方式を説明したものである。

第6図において、横軸は入力単語音声パターンであり、縦軸は単音節標準パターンを連結して作成された擬似単語標準パターンである。

いま、単語音声“アイチ(愛知)：a i tʃ i”が入力され、擬似単語標準パターン“アイチ(a i tʃ i)”とマッチングしたとき、そのDPバスは、図示の様に始点PからQ, R, S点を通り端点Tに終る経路をとり、“アイ(a i)”の端点U及び“チ(tʃ i)”の始点Vを通る経路、即ち理想的にマッチングが行われた場合の経路からはずれたものとなる。

この為、正しい照合が行われず、認識率が低下するという不都合を生じる。

#### 〔発明が解決しようとする問題点〕

従来の入力単語音声パターンを単音節標準バ

ーンから作成された擬似單語標準パターンと照合する單語音声認識方式は、前述の様に、マッチング時のD Pバスが理想的にマッチングが行われた場合のD Pバスからずれたものとなって正しい照合が行われない為、認識率が低下するという問題があつた。

本発明は、入力單語音声パターンを單音節標準パターンから作成された擬似單語標準パターンと照合して入力單語音声を認識する單語音声認識装置において、マッチング時のD Pバスが理想的にマッチングが行われた場合のD Pバスに合致する様にして認識率を向上させる様にした單語音声認識装置を提供することを目的とする。

#### 〔問題点を解決する為の手段〕

従来の入力單語音声パターンを單音節標準パターンから作成された擬似單語標準パターンと照合する單語音声認識方式では、マッチング時のD Pバスが理想的なD Pバスからずれるが、それは、入力單語音声パターンには、第6図に示す様に有

ことになる。

本発明は、従来の入力單語音声パターンを擬似單語標準パターンと照合する單語音声認識方式における誤認識の主な原因の1つが入力單語音声パターン中に存在する無音区間の及ぼす悪影響によるものであることに着目し、入力單語音声パターン中から無音区間パターンを除いて形成された圧縮單語音声パターンと擬似單語標準パターンとを照合させることにより、認識率を向上させる様にしたものである。

以下、従来の單語音声認識方式における前述の問題点を解決する為に本発明が講じた手段を、第1図を参照して説明する。

第1図は、本発明の基本構成をブロック図で示したものである。

第1図において、110は圧縮單語音声パターン作成手段で、入力單語音声より作成された入力單語音声パターンより無音区間パターンを取り除き、各有音区間パターンを詰めて圧縮單語音声パターンを作成する。

音区間「アイ (a i)」及び「チ (i ſ i)」の間に無音区間が介在しているのに対し、擬似單語標準パターンの「アイチ (a i i ſ i)」には、この様な無音区間が存在しないことに1つの大きな原因がある。

即ち、D P等の非線形伸縮を行って無音区間を含んだ入力單語音声パターンと無音区間を含まない擬似單語標準パターンとを照合する際、無理な対応付けが行われる為、マッチング時のD Pバスは理想的なD Pバスからずれ、正しい照合が行われることになる。そして、この傾向は、入力單語音声パターン中に占める無音区間の割合が多くなる程、顕著なものとなる。

又、入力單語音声パターンと擬似單語標準パターン間の累積距離中に占める無音区間パターンと單音節標準パターン間の距離の割合が増える程、無音区間パターン（雜音により、レベルは低いが多様なパターンを含んでいる）に出来るだけ近い單音節標準パターンを選ぶことになる為、累積距離誤差が大きくなつて、誤認識の可能性が高まる

120は擬似單語標準パターン作成手段で、各單音節標準パターンより認識対象となるカテゴリの單語群に属する各單語の擬似單語標準パターンを作成する。

130は單語認識手段で、圧縮單語音声パターン作成手段110より入力された圧縮單語音声パターンを擬似單語標準パターン作成手段120中の各擬似單語標準パターンと照合して入力單語音声を認識する。

#### 〔作用〕

入力單語音声から作成された入力單語音声パターンが入力されると、圧縮單語音声パターン作成手段110は、入力單語音声パターンより無音区間パターンを取り除き、各有音区間パターンを詰めて圧縮單語音声パターンを作成し、單語認識手段130に入力する。

一方、擬似單語標準パターン作成手段120には、各單音節標準パターンより認識対象となるカテゴリの單語群に属する各單語の擬似單語標準バ

ターンが予め作成されている。

単語認識手段130は、圧縮単語音声パターン作成手段110より入力された圧縮単語音声パターンを、擬似単語標準パターン作成手段120中の各擬似単語標準パターンと照合して、入力単語音声を認識する。この照合及び入力単語音声認識処理は、例えば、2段DP法により行うことがある。

以上の様にすることにより、入力単語音声パターン中に存在する無音区間パターンによる悪影響が除去されて擬似単語標準パターンとの照合が正しく行われ、入力単語音声の認識率を向上させることが出来る。

又、圧縮処理により照合対象のフレーム数が減少するので、単語認識時の照合処理量が低減され、圧縮単語音声パターン作成手段における処理量の増加があっても、全体の処理量を低減させることが出来る。

と入力単音節音声パターンに応じた切替えを行う。圧縮単語音声パターン作成手段110において、111は認識用区間検出部で、入力単語音声パターンから有音区間と無音区間の検出を行う。

112はパターン圧縮部で、認識用区間検出部111の検出した有音区間及び無音区間情報に基づいて、入力単語音声パターンより無音区間パターンを取り除き、各有音区間パターンを詰めて圧縮単語音声パターンを作成する。

擬似単語標準パターン作成手段120において、121は登録用区間検出部で、登録用の単音節音声パターンの区間検出を行って単音節標準パターンを作成する。

122は単音節標準パターン辞書で、作成された各単音節標準パターンが登録される。

123は単語辞書で、各単語の音節情報が格納されている。

124は擬似単語標準パターン作成部で、単語辞書123より認識対象となる単語群のカテゴリに属する各単語を取り出し、各単語の音節情報を

#### (実施例)

本発明の実施例を、第2図～第4図を参照して説明する。

第2図は本発明の一実施例の構成のブロック説明図、第3図は同実施例における区間検出方式の説明図、第4図は同実施例におけるDPマッチング方式の説明図である。

#### (A) 実施例の構成

第2図において、圧縮単語音声パターン作成手段110、擬似単語標準パターン作成手段120及び単語認識手段130については、第1図で説明した通りである。

140はマイクロホンで、話者(図示せず)の発声した単語音声又は単音節音声が入力される。

150はパラメタ抽出部で、マイクロホン140から入力された単語音声又は単音節音声の特徴を表すパラメタを抽出して、入力単語音声パターン又は入力単音節音声パターンを作成する。

160は切替え回路で、入力単語音声パターン

基づいて単音節標準パターン辞書122より所定の各単音節標準パターンを取り出し、各単語毎の擬似単語標準パターンを作成する。

単語認識手段130において、131はDP計算部で、パターン圧縮部112で作成された圧縮単語音声パターンを、擬似単語標準パターン作成部124で作成された各擬似単語標準パターンとDP照合を行い、各単語と圧縮単語音声パターンとの距離をそれぞれ算出する。

132は判定部で、DP計算部131により求められた各単語圧縮単語音声パターンとの距離を比較し、その距離の最も小さい単語を認識単語と判定する。

#### (B) 実施例の動作

実施例の動作を、第3図及び第4図を参照し、各動作に分けて説明する。

##### (B-1) 登録動作

話者の発声した単語音声に対する認識処理が行

われる前に、単音節標準パターン辞書122には各単音節の標準パターンが登録され、更に、擬似単語標準パターンが作成される。

単音節標準パターン辞書122に各単音節標準パターンを登録する場合は、切替え回路160を登録用区間検出部121側に接続し、マイクロホン140より単音節音声をバラメタ抽出部150に入力する。

バラメタ抽出部150は、入力された単音節音声の特徴を表すバラメタを抽出して、入力単音節音声パターンSPを作成する。

作成された単音節音声パターンSPは、特徴ベクトルの時系列であり、各特徴ベクトルは、q個(例えば16個)の帯域フィルタのパワースペクトルをq次のベクトル量で表したものである。従って、横軸に時間tをとり、縦軸にパワーをとると、入力単音節パターンSPは、第3図(a)に示す様なパターンを形成する。

この入力単音節音声パターンSPに対し、2種類の閾値h<sub>1</sub>及びh<sub>2</sub>を設ける。閾値h<sub>1</sub>は、雑

音レベルよりは高く、各入力単音節音声パターンのパワーの最大値の中で最も低い値の近傍に選定される。h<sub>2</sub>は雑音レベル、即ち無音区間パターンのパワーレベルの最大値の近傍に選定される。

登録用区間検出部121は、入力待ちになってから、入力単音節音声パターンのパワーが閾値h<sub>1</sub>を初めて越えたフレーム(f<sub>1</sub>)を探し、このフレームf<sub>1</sub>から両側でパワーが閾値h<sub>2</sub>以上である連続した区間(始端f<sub>1</sub>～終端f<sub>2</sub>)を単音節標準パターンの音声区間として検出する(第3図(b)参照)。

これにより、雑音N<sub>1</sub>～N<sub>3</sub>を除いた、始端f<sub>1</sub>から終端f<sub>2</sub>間の入力単音節音声パターン部分が登録用の単音節標準パターンとして抽出されて、単音節標準パターン辞書122に登録される。

認識対象となる単語群のカテゴリが決まると、擬似単語標準パターン作成部124は、単語辞書123より認識対象となる単語群のカテゴリに属する各単語を取り出し、各単語の音節情報に基づいて単音節標準パターン辞書122より所定の各

単音節標準パターンを取り出し、各単語毎の擬似単語標準パターンを作成する。

#### (B-2) 圧縮単語音声パターン作成動作

入力された単語音声パターンに対する認識処理を行う場合は、切替え回路160を認識用区間検出部111側に接続して、圧縮単語音声パターンの作成が行われる。

マイクロホン140より未知単語音声が入力されると、前述の単音節標準パターンの登録の場合と同様にして、バラメタ抽出部150は、入力単語音声パターンWPを作成して認識用区間検出部111に入力する。

作成された入力単語音声パターンWPは、入力単音節音声パターンと同様な特徴ベクトルの時系列であり、各特徴ベクトルはq個の帯域フィルタのパワースペクトルをq次のベクトル量で表したものである。従って、横軸に時間tをとり、縦軸にパワーをとると、入力単語音声パターンWPは、第3図(c)に示す様なパターンを形成する。

この入力単語音声パターンWPに対し、前述の登録用区間検出部121の場合と同様な閾値h<sub>1</sub>及びh<sub>2</sub>が設定される(第3図(c)参照)。

認識用区間検出部111は、入力待ちになってから、入力単語音声パターンWPのパワーが閾値h<sub>1</sub>を初めて越えたフレーム(f<sub>1</sub>)を探し、このフレームf<sub>1</sub>から両側でパワーが閾値h<sub>2</sub>以上の区間(始端f<sub>1</sub>～f<sub>2</sub>、f<sub>2</sub>～f<sub>3</sub>、f<sub>3</sub>～f<sub>4</sub>、f<sub>4</sub>～f<sub>5</sub>)を探す。その際、閾値h<sub>2</sub>以下になる区間(f<sub>1</sub>～f<sub>2</sub>、f<sub>2</sub>～f<sub>3</sub>)が所定の長さし。より小さいときは、無音区間として入力単語音声パターンに含ませ、L<sub>1</sub>を越えた場合(例えばf<sub>1</sub>～f<sub>2</sub>、f<sub>2</sub>～f<sub>3</sub>)は、雑音として無視する。L<sub>1</sub>は、各単語音声中に含まれる各無音区間中の最大値に基づいて選定される。

これにより、始端f<sub>1</sub>から終端f<sub>5</sub>間の入力単語音声パターン部分が、圧縮の対象となる入力単語音声パターンWPcとして抽出される。

認識用区間検出部111は、更に、この圧縮の対象となる入力単語音声パターンWPcにおいて、

そのパワーレベルが閾値  $h_1$  以上である区間、即ち有音区間 ( $t_1 \sim t_2$ ,  $t_3 \sim t_4$ ,  $t_5 \sim t_6$ ) と閾値  $h_2$  より低い区間、即ち無音区間 ( $t_2 \sim t_3$ ,  $t_4 \sim t_5$ ) を検出する(第3図(b)参照)。

パターン圧縮部112は、認識用区間検出部111の検出した有音区間及び無音区間情報に基づいて、圧縮対象となる入力単語音声パターンWPより無音区間 ( $t_2 \sim t_3$ ,  $t_4 \sim t_5$ ) のパターンを取り除き、各有音区間 ( $t_1 \sim t_2$ ,  $t_3 \sim t_4$ ,  $t_5 \sim t_6$ ) の各パターンを詰めて、圧縮単語音声パターンを作成する。

第4図の横軸に示されているパターンは、第6図の横軸に示されている入力単語音声パターン“アイチ (a i t f i) ”を圧縮して得られた圧縮単語音声パターン“アイチ (a i t f i) ”を示したものである。

#### (B-3) 単語認識動作

DP計算部131は、パターン圧縮部112で

作成された圧縮単語音声パターンを、擬似単語標準パターン作成部124で作成された各擬似単語標準パターンとDP照合を行い、各単語と圧縮単語音声パターンとの距離をそれぞれ算出する。

第4図は、第6図に示す入力単語音声パターン“アイチ (a i t f i) ”を圧縮して得られた圧縮単語音声パターン“アイチ (a i t f i) ”を各擬似単語標準パターンと照合を行い、擬似単語標準パターン“アイチ (a i t f i) ”とマッチングした場合を示したものである。

図示の様に、圧縮単語音声パターンの“アイ (a i) ”及び“チ (t f i) ”の部分は、擬似単語標準パターンの“アイ (a i) ”及び“チ (t f i) ”の部分と、正しいDPバスP'～R'及びR'～T'によってそれぞれ照合が行われる。従って、擬似単語標準パターン“アイチ (a i t f i) ”と単語“アイチ(愛知) ”の距離が最も小さい値をもって算出されることになる。

判定部132は、DP計算部131により算出

された各単語と圧縮単語音声パターンとの距離を比較し、その距離の最も小さい単語、即ち“アイチ(愛知) ”を認識単語と判定する。

以上、本発明の一実施例について説明したが、単音節標準パターン辞書の代りに、音素単位で登録された標準パターンを用いることも出来る。その場合の各演算処理は、前述の単音節単位で登録されている場合と同様にして行うことが出来る。

#### (発明の効果)

以上説明した様に、本発明によれば、次の諸効果が得られる。

(イ) 入力単語音声パターン中に存在する無音区間の影響によるDPバスのそれが除去されて正しい照合が行われるので、認識率を向上させることが出来る。

(ロ) 圧縮処理により照合対象のフレーム数が減少するので、単語認識時のDP照合計算処理量が低減され、圧縮単語音声パターン作成手段における処理量の増加があっても、全体の処理量

を低減させることができる。

#### 4. 図面の簡単な説明

第1図…本発明の基本構成の説明図、

第2図…本発明の一実施例の構成の説明図、

第3図…同実施例における区間検出方式の説明図、

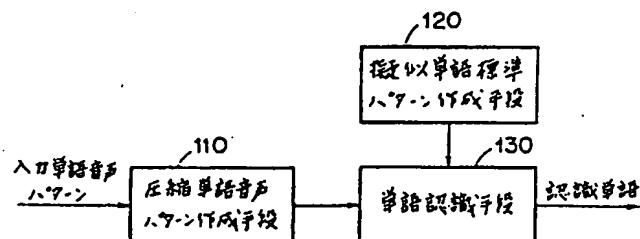
第4図…同実施例におけるDPマッチング方式の説明図、

第5図…従来の擬似単語標準パターンによる単語音声認識方式の説明図、

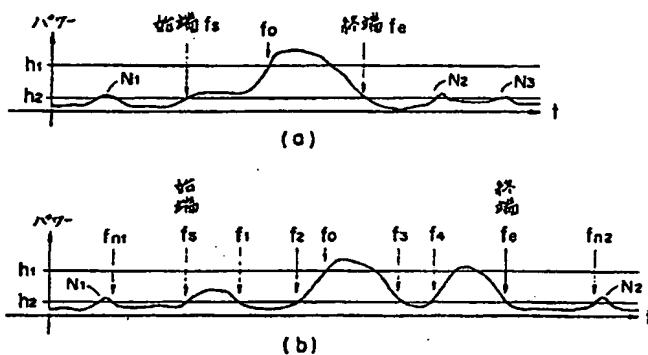
第6図…従来の擬似単語標準パターンによる単語音声認識方式におけるDPマッチング方式の説明図。

第1図及び第2図において、

110…圧縮単語音声パターン作成手段、120…擬似単語標準パターン作成手段、130…単語認識手段、140…マイクロホン、150…パラメタ抽出部、160…切替え回路。

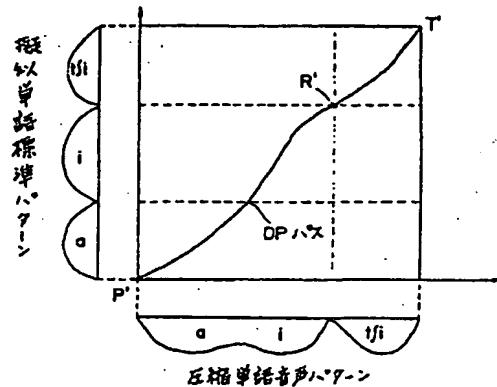


第1図



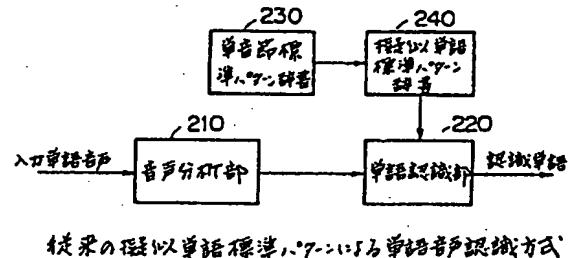
実施例1: ハーフーン区間検出方式

第3図

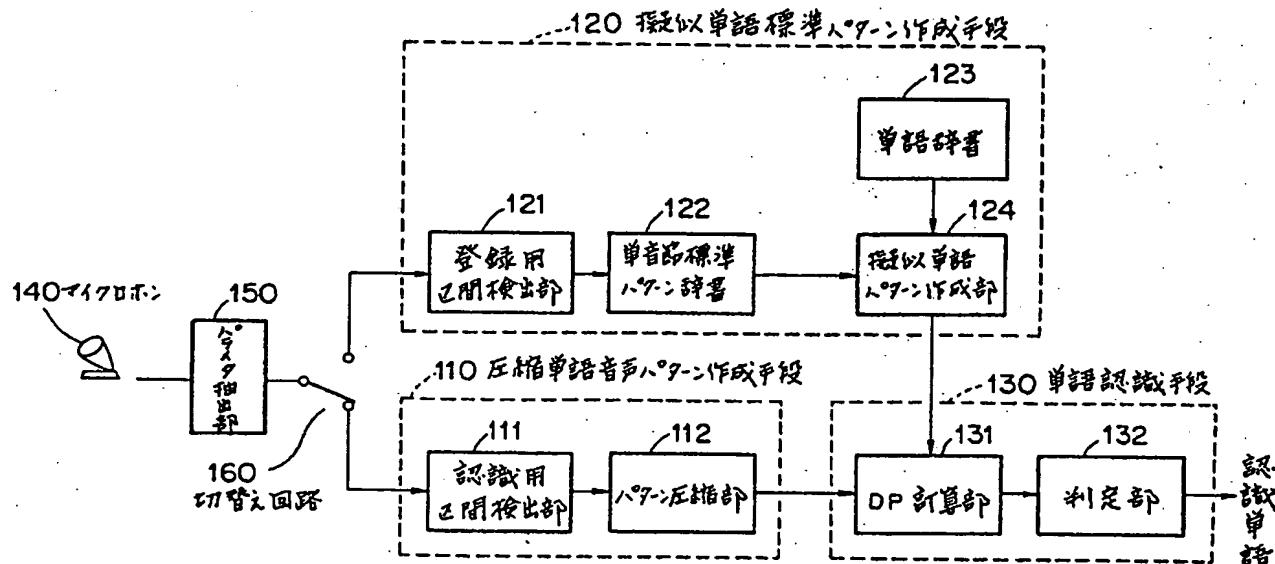


実施例1: ハーフーン DP マッチング方式

第4図

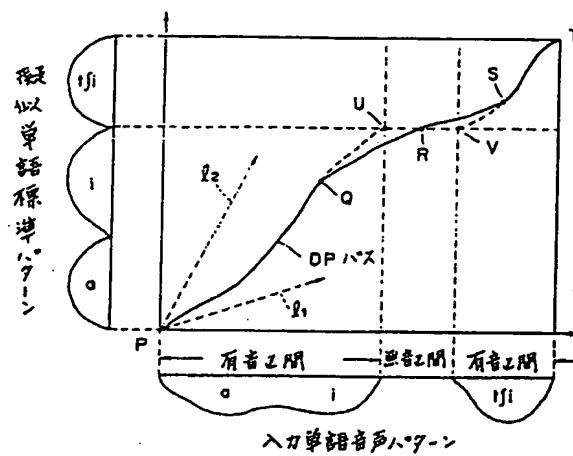


第5図



実施例の構成

第2図



従来の擬似單語標準パターンによる單語  
音声認識方式におけるDPマッチング方式

第6図